1. The curves for each designated value of “mtry”, when plotted as a function of “ntree”, appeared to behave similarly. All of them were very coarse and spiky, with a sharp decrease from around the range where “ntree” was between 1 and 100, with the overall slope of the curve of best fit stagnating and becoming very horizontal afterward. This seems to suggest that a number of trees around 100 provides a sufficiently accurate random forest model on the Boston dataset.
2. Sales
   1. Split
   2. The tree appears to depend most heavily on whether or not the shelving location of the carseat is good or not. The carseats with the most sales were the ones with both good locations and a price less than $135.5. The carseats with the least sales were the ones with bad or middling locations, with prices between $92 and $132.5, and with a competing carseat’s price of less than $123.5. The resulting test MSE from this regression tree is about 4.47.
   3. The test MSE does not appear to benefit from pruning the tree, as the optimal size is the largest available one, which is the same size as the original tree.
   4. Bagging MSE is significantly better than of using a single regression tree, with a value of 2.54. The most significant variables are, in descending order, price, shelving location, competitor price, and advertisement. The rest all appear to have no real impact on the value of sales.
   5. The test MSE of using default value random forests is 3.30, somewhat worse than the bagging MSE. The most significant variables are the same as in the bagging model, with no real change in the order. For this dataset, the more variables considered, the lower the test error rate obtained.
3. OJ
   1. Split
   2. Training error rate is 12.65%, and there are 17 terminal nodes. The mean deviance doesn’t appear to be too high, only 0.573. The variables included in the tree are “LoyalCH”, “StoreID”, “DiscMM”, “DiscCH”, “WeekofPurchase”, “SalePriceMM”, “Store7”, and “PriceDiff”.
   3. In the first split, the criterion is whether or not LoyalCH<0.51. 96 observations satisfy this criterion, with a deviance of 122.2. The overall prediction for observations falling under this branch is MM, with a third of the observations predicting CH and two-thirds predicting MM.
   4. The tree’s first split based on the value of LoyalCH is extremely prominent in terms of impacting the classification of observations. All but 2 of the leaf nodes on the left side (LoyalCH<0.51) will end in a prediction of MM for the observation, while all but 2 of the leaf nodes on the right side (LoyalCH>0.51) will predict CH for the observation.
   5. Test
   6. CV
   7. Plot CV
   8. Lowest CV error comes from a tree of size 2.
   9. Pruned tree created
   10. The training error of the unpruned tree is lower than that of the pruned tree.
   11. The test error of the unpruned tree is slightly higher than that of the pruned tree.
4. Salary from Hitters
   1. Removed Observations
   2. Split
   3. Boosting and training MSE
   4. Test MSE
   5. Compared to both the lasso and standard linear regression, boosting on the Hitters dataset in this case boasts a much lower test MSE (about 50% of the former two) than either of the regression approaches tested.
   6. The most significant predictors appear to be CHits, CRuns, and CRBI, followed after a steep drop in relative influence by CHmRun, Walks, and Putouts.
   7. The bagging approach provides a test MSE of 0.232, which is significantly lower than that of the lasso and standard linear regression, and is about the same as the boosting approach’s test MSE of 0.243.
5. Caravan
   1. Split
   2. The most important predictors appear to be MBERMIDD, MAUT2, and PPERSAUT.
   3. About 12.1% of the people that were predicted to make a purchase by the boosting model actually made one. This is slightly worse than the logistic regression’s performance, which had 14% of the people predicted to make a purchase actually making one.
6. To be done in the future